(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/012695 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/US02/24054

(22) International Filing Date: 31 July 2002 (31.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/308,594 31 July 2001 (31.07.2001) US

(71) Applicant: **GRACENOTE, INC.** [US/US]; 2141 4th Street, Berkeley, CA 94710 (US).

(72) Inventors: **ROBERTS, Dale, T.**; 15 Oak Springs Drive, San Anselmo, CA 94960 (US). **HYMAN, David**; 285 Los Altos Drive, Kensington, CA 94708 (US). **WHITE, Stephen**; 2326 McKinley Avenue, Berkeley, CA 94703 (US).

(74) Agent: **GOLLHOFER, Richard, A.**; Staas & Halsey LLP, Suite 500, 700 Eleventh Street, N.W., Washington, DC 20001 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/012695 A2

(54) Title: MULTIPLE STEP IDENTIFICATION OF RECORDINGS

(57) Abstract: Multiple information is extracted from an unknown recording and information associated therewith. Associated information includes the filename, if the recording is a computer file in, e.g., MP3 format, or table of contents (TOC) data, if the recording is on a removable medium, such as a compact disc. At least one and preferably several algorithmically determined fingerprints are extracted from the recording using one or more fingerprint extraction methods. The information extracted is compared with corresponding information in a database maintained for reference recordings. Identification starts with the most accurate and efficient method available, e.g., using a hash ID, a unique ID or text. Fingerprint matching is used to confirm other matches and validation is performed by comparing the duration of the unknown and a possibly matching reference recording.

TITLE OF THE INVENTION

MULTIPLE STEP IDENTIFICATION OF RECORDINGS

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application is related to and claims priority to U.S. provisional application entitled DIGITAL MUSIC MULTIPLE STEP IDENTIFICATION METHOD AND SYSTEM having serial number 60/308,594, by Dale T. Roberts, et al., filed July 31, 2001, and incorporated by reference herein.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0002] The present invention is directed to recognition of recordings from their content, and, more particularly to combining fingerprint recognition with other information about a recording to increase reliability of recognition and to accomplish reliable recognition efficiently by using the least expensive forms of recognition first and layering on more complex forms as needed.

2. Description of the Related Art

[0003] There are many uses for recognition of audio (and video) recordings. Many of the uses relate to compensation or control by the rights holders for reproduction and performance of the works recorded. This use of such systems has increased in importance since the development of file sharing software, such as Napster, and the many other similar services available at the end of the twentieth century and the beginning of the twenty first century. Although the need for accurate recognition has been significant for several years, no system has been successful in meeting this need.

[0004] Another use of recording recognition is to provide added value to users when listening (or watching) recordings. One example is the CDDB Music Recognition Service from Gracenote, Inc. of Berkeley, California which recognizes compact discs (CDs) and supplies information regarding a recognized CD, such as album name, artist, track names and access to related content on the Internet, including album covers, artist and fan websites, etc. While the CDDB service is effective for recognizing compact discs, there are several draw backs in using it to recognize files that are not stored on a removable disc, such as CD or DVD.

[0005] All audio fingerprinting techniques have "blind spots", places where a system using that technique sees similarities and differences in audio where it shouldn't. By relying on just one fingerprinting technique, single source solutions are less accurate when encountering a 'blind spot'.

[0006] One of the more popular uses for the Gracenote CDDB system is in applications that digitally encode audio files into MP3 and other formats. These encoding applications utilize Gracenote's CDDB service to recognize the compact disc being encoded and to write the correct metadata into the title and ID tags. Gracenote's CDDB service returns a unique ID (TUID) for each track and supports the insertion of such IDs in the ID3V2 tags for MP3 files. The TUID is both hashed and proprietary, and can only be read by the Gracenote system. However, the ID3V2 tags can easily be manipulated to store a TUID for one file in the ID3V2 tag for another file and therefore, the TUID alone is not a reliable identifier of the audio content in a file.

[0007] Gracenote's CDDB service also provides text matching capability that can be utilized to identify digital audio files from their file names, file paths, ID tags (titles), etc. by matching the text extracted by a client device to a metadata database of track, artist, and album names. Although this text matching utilizes user-generated spelling variants associated with each record to improve recognition, there has been no way to verify that the text matches the audio content of the recording once the recording has been separated from a compact disc and stored in a file in any format.

SUMMARY OF THE INVENTION

[0008] An aspect of the present invention is maximizing identification of recordings while minimizing resource usage.

[0009] Another aspect of the present invention is using multiple identification methods so that resource intensive methods, such as audio fingerprinting, are employed only when necessary.

[0010] A further aspect of the invention is minimization of processing of unidentified data.

[0011] Yet another aspect of the present invention is to use the least expensive recognition technique, with progressively more expensive recognition techniques layered onto the process until a desired confidence level is reached.

[0012] A still further aspect of the invention is validation of content-based identification of a recording by comparing text associated with an unidentified recording and text associated with identification records.

[0013] Yet another aspect of the present invention is use of recording identification methods from different sources to increase reliability.

[0014] A still further aspect of the invention is validation of content-based recording identification using fuzzy track length analysis.

[0015] Yet another aspect of the invention is automatic extraction of identification data for use in a reference database and for identification of recordings.

[0016] A still further aspect of the invention is that unidentified recordings are periodically re-run through the system to determine if recently added data or recently improved techniques will result in recognition.

[0017] The above aspects can be attained by a method of identifying recordings by extracting information about an unknown recording stored in media possessed by a user and at least one algorithmically determined fingerprint from at least one portion of the unknown recording; determining a possible identification of the unknown recording using at least one piece of the information extracted from the unknown recording and an identification database of corresponding information for reference recordings; and identifying the unknown recording when the possible identification based on each of the at least one piece of the information in combination with the at least one algorithmically determined fingerprint identifies a single reference recording with respective confidence levels. The at least one portion of the unknown recording may contain audio, video or both.

[0018] Preferably, the database is maintained by a provider of identification services which supplies unique identifiers that can be recognized only by servers under the control of the provider of identification services. The unique identifiers are associated with recordings once they have been identified. Subsequently, copies of the recordings are recognized using the unique identifiers to greatly speed up the process. The unique identifiers optionally are cached in high-speed RAM or specially indexed database tables.

[0019] When non-waveform data is not available for an unknown recording, the unknown recording is preferably identified by extracting fingerprints from at least one portion of the unknown recording using a plurality of algorithms; determining a possible identification of the unknown recording using at least two of the fingerprints extracted from the unknown recording and at least one database of correspondingly generated fingerprints for reference recordings; and identifying the unknown recording when the possible identification based on each of the fingerprints identifies a single reference recording with respective confidence levels.

[0020] Preferably, an existing database, used to identify recordings possessed by users, which does not contain fingerprint information is expanded by obtaining non-waveform data associated with a recording possessed by a user of the database; extracting at least one fingerprint from at least one portion of the recording; and storing the at least one fingerprint as identifying

information for the recording, when a match is found in the database for the non-waveform data. One example is that during the process of encoding digital music files from an audio CD possessed by a user, a recognition system can be used to identify the audio CD so that fingerprints extracted during the encoding process can be directly associated with the audio CD using a unique ID system.

[0021] Recognition of recordings using either fingerprints or unique identifiers is preferably validated by other information maintained in the identification database, such as the length of the recording or a numeric identifier embedded within the recording. Information about recordings that do not pass validation or match some, but not all of the information used for identification, may be stored for later analysis of the reason for the error. If the fingerprints are obtained as described above, there may have been an error in obtaining the fingerprint. Therefore, errors may be output to an operator, or the system could correct the information stored in the database, based on recognition of patterns in the information that is stored for improper matches. For example, if a large percentage of matching fingerprints are stored, but the other information consistently does not match them, there could be an error in the fingerprint database which needs to be flagged to an operator.

[0022] The present invention includes a system for identifying recordings that includes an extraction unit to extract information about an unknown recording stored in media possessed by a user and at least one algorithmically determined fingerprint from at least one portion of the unknown recording; and an identification unit, coupled to the extraction unit, to make a possible identification of the unknown recording using at least one piece of the information extracted from the unknown recording and an identification database of corresponding information for reference recordings, and to identify the unknown recording when the possible identification based on each of the at least one piece of the information in combination with the at least one algorithmically determined fingerprint identifies a single reference recording with respective confidence levels.

[0023] The present invention also includes a system for identifying recordings that includes an extraction unit to extract fingerprints from at least one portion of an unknown recording using a plurality of algorithms, and an identification unit, coupled to said extraction unit, to make a possible identification of the unknown recording using at least two of the fingerprints extracted from the unknown recording and at least one database of correspondingly generated fingerprints for reference recordings, and to identify the unknown recording when the possible identification

based on each of the fingerprints identifies a single reference recording with respective confidence levels.

[0024] In either of the systems described above, the extraction unit is typically a client unit connected by a network, such as the Internet, to at least one server as the identification unit. The client device may be a personal computer with a drive accessing the recording, a consumer electronics device with a network connection, or a server computer transmitting the unknown recording from one location to another. Furthermore, a portion of the database may be available locally and the extraction unit and identification unit may reside in the same device and share components.

[0025] The present invention also includes a system for obtaining reference information stored in a database used to identify unknown recordings, including a receiving unit to obtain non-waveform data associated with a recording possessed by a user of the database for identification of recordings possessed by the user; an extraction unit to extract at least one fingerprint from at least one portion of the recording; and a storage unit, coupled to said receiving unit and said extraction unit, to store the at least one fingerprint as identifying information for the recording, when a match is found in the database for the non-waveform data.

[0026] These together with other aspects and advantages which will be subsequently apparent, reside in the details of construction and operation as more fully hereinafter described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals refer to like parts throughout.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a functional block diagram of a system according to the present invention.

Figure 2 is flowchart of a fingerprint extraction according to the present invention.

Figure 3 is a flowchart of a method of recognizing unknown recordings.

Figures 4A-4C are a block diagram of a system according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0027] According to the present invention, a suite of identification components are provided in a system like that illustrated in Fig. 1 to facilitate analysis and identification of audio (and video) files utilizing multiple methods. Preferably, an existing database 90 containing recording identifiers and text data is combined with text-based digital audio and audio fingerprinting identification methods. Preferably, the text data in database 90 is obtained from user

submissions and includes user-submitted spelling variants. One such database is available as the CDDB Music Recognition Service from Gracenote, Inc..

[0028] As illustrated in Fig. 1, a recording 100 is accessed by client device 110 via any conventional method, such as reading a digital audio file from a hard drive or a compact disc. Information is extracted from recording 100 and associated information (metadata). Fingerprints are extracted from recording 100, as described in more detail below. The information that is extracted from the metadata includes the duration of the recording which is the track length (from the TOC) for a CD track, the filename and ID3 tag if the recording is in an MP3 file, and the table of contents (TOC) data if the recording is on a CD. If the file containing the recording was produced by a client device operating according to the invention, a unique ID will be extracted from the ID3 file, but initially it will be assumed that information is not available.

[0029] In an exemplary embodiment, the extracted information is sent from client 110 to server 120 to determine a possible identification of the unknown recording using at least one piece of the information extracted from recording 100 and a database 130 of correspondingly generated fingerprints for reference recordings. If text or a unique ID were extracted, an attempt is made to find a match. If a match is found using the text or unique ID, at least one algorithmically determined fingerprint is compared with the fingerprint(s) stored in the matching records to determine whether there is a single reference recording that matches the information extracted from recording 100 with respective confidence levels for each item of information that matches. If no matches can be found based on text and unique ID, an attempt is made to identify the a single reference recording using at least two of the fingerprints extracted from recording 100. If a single reference recording is located using either method, preferably the duration of recording 100 is compared with the duration of the single reference recording as a final validation step.

[0030] Preferably related metadata is used for validation of the match obtained by fingerprint recognition. Like any recognition system fingerprinting can produce erroneous results. Without a validation component such an error can propagate throughout the system and return erroneous data to large percentages of users. The use of validation criteria such as track length comparison enables the system to catch potential errors and flag them for validation.

[0031] A system according to the present invention preferably includes custom result reporting and flexible administrative interfaces 130 to enable weighting of various identification methods and the order of their engagement. Analysis of successful match rates for specific identification methods allows an administrator to manipulate the identifying criteria for each component to

maximize the identification probability. A system according to the present invention preferably incorporates usage data from over 28 million users utilizing the CDDB database via Gracenote Data Services division, to help guide results 140.

[0032] The flexibility of a system according to the present invention allows different configurations to be used for identifying recordings in different environments. An application that monitors streaming audio, for example, requires a very different system and solution architecture than one that identifies files in a peer-to-peer system, or one that identifies analog input. However the present invention can be configured for identification of recordings in each of these situations.

[0033] A system according to the present invention maximizes identification while minimizing resource usage. The use of multiple identification methods ensures that more resource intensive methods, such as audio fingerprinting are employed only when necessary. The use of multiple audio fingerprinting technologies reduces data collision and covers any "blind spots" in a given audio fingerprint technology. The "blind spots" found in single source fingerprinting systems, are avoided by using multiple sources for different fingerprinting techniques. This also provides the ability to fine tune deployment for specific target applications.

[0034] Preferably, fingerprints are obtained using multiple fingerprint recognition services using the method illustrated in Fig. 2. This increases the ability of the system to accurately recognize recordings of various types.

[0035] As illustrated in Fig. 2, when unidentified (unknown) recording 100 is accessed by fingerprint extraction client 110, if possible conventional TOC/file recognition is performed by recognition system 210 and results 220 are returned to fingerprint client 110. Results 220 include a unique identifier (TUID) that points into a master metadata database (not shown in Fig. 2), if the TUID is found. Recording 100 is also processed by fingerprint extractor 230 using at least one and preferably several different algorithmically derived fingerprint extraction systems to obtain fingerprint(s) which are stored in fingerprint/ID send cache 240. As described below in more detail, instructions are received regarding when fingerprint uploader 250 should send the fingerprints to fingerprint recognition server 120.

[0036] In fingerprint recognition server 120, the fingerprints transmitted by fingerprint uploader 250 are initially stored in fingerprint receive cache 260. The fingerprints then undergo fingerprint validation 270 using an algorithmic comparator that attempts to cross-correlate fingerprints for a recording with fingerprints uploaded and extracted by different end users. If it

is found that the fingerprints are substantially similar, they would be validated. This is not the only method that's available for validation, but serves as one example of a process that could be used to reject bad data.

[0037] In this embodiment, fingerprints that are determined to be valid and related undergo stitching 280. For example, if fingerprints are taken from 30 second segments of the recording, the fingerprints are assembled into a continuous fingerprint stream. This could simplify recognition of segments of the recording. The resulting fingerprints are stored in fingerprint database 290 associated with existing database 90 (Fig. 1).

[0038] The CDDB database has in part been generated through user submissions to create a metadata database with over 12 million tracks and 900,000 albums as of mid-2002. This database contains both basic metadata (artist, album, and track names) as well as extended data (genre, label, etc.).

[0039] A similar distributed collection method may be utilized in the creation of a waveform database using the system illustrated in Fig. 1. In the case where recording 100 is a raw audio waveform, e.g., when a CD is encoded into another format, such as an MP3 file, client device 110 obtains non-waveform data associated with recording 100 which is possessed by a user of database 90 and executes extraction algorithm(s) to extract fingerprints from at least one portion of the recording. The fingerprints are then sent to server 120 with a unique ID, preferably derived from the TOC of the CD. When the unique ID is available, i.e., , when a match is found in the database for the non-waveform data, server 120 is able to associate the appropriate metadata in database 90 and the fingerprint(s) with same level of accuracy as identification of CDs by the existing database 90 which is provided for identification of recordings possessed by users. Fingerprints dynamically gathered in this manner may be sent to a fingerprint collection server (not shown in Fig. 1) which would accumulate fingerprints from authenticated clients, as described in more detail below, prior to storing the at least one fingerprint as identifying information for the recording.

[0040] Multiple fingerprint gathering extractors can also be run over a set of static waveforms from a commercial encoder such as Loudeye or Muse. The challenge with this approach is associating the fingerprints with the appropriate metadata. The method described above enables audio fingerprints to be logically associated with parent records and associated back to the original audio source. In the preferred embodiment, the unique ID provides differentiation

between live and studio versions of the same song while simultaneously linking those records to the same artist and their respective albums.

[0041] Preferably server(s) 120 store information in a parallel record set that are linked with unique IDs. When client 110 asks server 120 to recognize media (CD, digital audio file, video file) server 120 may also return a record about how fingerprints should be gathered for this particular CD. This is called the Gathering Instructions Record (GIR). The GIR may include a set of instructions that the remote fingerprint gathering code follows. The record may be pre-computed in off hours or may be dynamically computed at the time of recognition.

[0042] Server 120 may use information it knows about the popularity of a CD to drive decisions about gathering. Everything about a rare CD could be gathered, because the opportunity to get the fingerprints would not want to be missed (even if it was somewhat burdensome to the user). The opposite situation could be true for a very popular CD. The load may be distributed across many users so that they would not even notice that any work for fingerprint gathering was occurring.

[0043] The rules and procedures for building the GIR may be manual, automated and may change over time. They may also be applied uniquely to specific users, applications or geographic locations.

[0044] In one embodiment, the server dynamically gathers fingerprints by modifying the GIR to remove fingerprints that have been gathered previously. The frequency of updating GIRs may vary from instant to delays of days, weeks or months. Some example instructions that may be included in the GIR are:

- A list of track and segments to be gathered and their priority.
- A fingerprint generator algorithm to use.
- Parameters that tell the fingerprint generator how to process the fingerprint, such as:
  - Frequency of audio samples
  - Bands of the frequency domain to process
  - Resolution of the fingerprint
  - Desired Quality of Audio
- When to do the fingerprint gathering, such as
  - Before encoding the track
  - After encoding the track
  - In parallel with encoding the track

- Instructions for caching the fingerprint and when to transmit it back to the server, such as
  - Before encoding the track
  - After encoding the track
  - After the CD has been fully encoded
  - When the communication channel back to the server is not busy
  - When the next CD is looked up
  - When a group of fingerprints is ready for transmission
- Instructions to take CPU power into the process so as to not overload the computer

[0045] Preferably, the system attempts to improve the quality of the fingerprints during operation. Quality of the source signal, the parameters used for fingerprinting, along with improvements in the fingerprinting algorithms will result in a complex quality matrix that is used by server 120 to determine what fingerprints to gather if higher quality is available. An example of source quality is provided below: Preferably, database 90 or a similar database maintained by fingerprint collection server(s) stores the source quality for fingerprints stored in the database, so that when a fingerprint from higher quality source is available, the fingerprint may be replaced.

Source Quality Table

Name	Bit Rate	Compression	Error Correction	Quality Index
CD_Audio_HEC	44100kbps	None	Hardware	1
CD_Audio_SEC	44100kbps	None	Software	2
CD_Audio	44100kbps	None	None	3
CDR_Audio	44100kbps	None	None	4
CDR_Made_From_MP3	44100kbps	mp3	None	5
MP3_File	160kbps	mp3	None	6

[0046] Fingerprints dynamically gathered may contain information that helps validate quality. Information such as errors while reading from the media may be sent up to the fingerprint collector. The system may reject fingerprints that had high error rates from the source media.

[0047] As noted above, instead of immediately storing a fingerprint, multiple fingerprints for a recording may be gathered in by a fingerprint collection server prior to being added to the database. These fingerprints may be compared algorithmically to determine their correlation. If correlation is not adequate then additional fingerprints may be gathered until adequate correlation is achieved and one of the fingerprints or a composite fingerprint is stored in the database. This prevents bad fingerprints from becoming part of the database.

[0048] Stitching of the segmented fingerprints may be necessary since slight variations in timing could result in overlap of the fingerprints. Algorithmic stitching could result in a higher quality continuous fingerprint. Simple stitching appends segmented fingerprints in order of appearance in the recording. Complex stitching could involve scaling different qualities of fingerprints to the lowest common denominator and then appending them in order of their appearance in the recording. Preferably some form of mathematical fitting is utilized if the fingerprint segmentation contains jitter, so that appending is a fuzzy process rather simple addition of the datastream.

[0049] One example of audio fingerprinting that can be used is described in the U.S. patent application entitled Automatic Identification of Sound Recordings, filed by Maxwell Wells et al. on July 22, 2002 and incorporated herein by reference. However, any known algorithmically derived fingerprinting technique may be used, not only for digital audio, but also video, TV programs (both analog and digital) and DVDs. Appropriate identifiers and recognition techniques will be used for the media to be recognized in a particular application.

[0050] The present invention provides great flexibility and can be utilized for a wide variety of environments, including MP3 recognition in a peer-to-peer environment, or identification of an audio stream for monitoring and reporting purposes. No other solution is known to use multiple recognition components; so it is the only solution that can be customized to meet the needs of any audio (or video) recognition application.

[0051] A functional description for a deployment of the present invention in a peer-to-peer application will be described below with reference to Fig. 3. In this embodiment, audio files are identified before providing public access to them, to determine if the files are allowed in the system, a process known as "filter-in".

[0052] Client device 110 (Fig. 1) extracts information 310 (Fig. 3) from an audio file at the time of upload to server 120 (Fig. 1). The extracted information preferably includes non-waveform data, such as a unique ID, ID3 tag, filename text data, track duration, etc. and fingerprint(s) extracted from the recording and sent to server 120 for recognition.

[0053] The initial match 320 is performed against the unique ID, if present. Use of Gracenote's TUID enables a match to be returned with 99.9% accuracy. This is also the least resource intensive recognition method and can achieve very fast recognition rates. If the unique ID is present the system moves to the validation stage. If no unique ID is present the system attempts identification using the next recognition methods 330.

[0054] In this embodiment, text-based identification is tried next, using a metadata database, such as the Gracenote CDDB service which contains over 900,000 albums and over 12 million songs. Text matching utilizes available text, such as the filename, file path or text within the ID3 tag for MP3 files, to provide a set of data from which to attempt recognition. If an acceptable match is returned, the system moves to the validation stage. If a successful match is not returned, the system attempts identification utilizing the next recognition method.

[0055] The next step is fingerprint identification, in this case using audio fingerprints. The fingerprints from an unknown recording are compared to the fingerprints in database 90 for reference recordings, one fingerprint at a time (or in parallel using different processors for different fingerprints). Each fingerprinting technology returns a match and a level of confidence. If a single reference recording has acceptable confidence levels the system moves to the validation stage. If an unsuccessful match is returned the system can, depending on the target application, ask the user for validation of the most likely result or it can return a "no match found" result.

[0056] Validation is a key component to any successful recognition system. Preferably, key file attributes such as the duration of the recoding, are used to validate that a file is what the recognition system says it is by comparing an extracted length of the unknown recording with a stored length of the single reference recording.

[0057] Preferably heuristic and voting algorithms 340 are used to determine if a match is what the system says it is. This self-monitoring reduces the possibility that the system returns inaccurate data that pollutes the system. The heuristics may be manually controlled or algorithmically controlled to produce the best match. These heuristics may also be used to determine which recognition techniques to apply and in what sequence.

[0058] The administrator of each application can determine the level of accuracy needed by each stage (or component) of the system, and therefore has explicit control in optimizing the system. For example, if a 90% aggregate match is required the system administrator can use administrative interfaces 130 to adjust the levels of acceptable return to 90% and a successful result will not be generated unless that threshold is met. The administrator can also set result levels for each component. For example, a 99% text match can be required but only an 85% audio fingerprint match.

[0059] Once a successful identification is returned the file will be retagged 350 with the unique ID allowing for population of the file with the correct ID throughout the system. As a result, future identification of the file will require the least resource intensive recognition method.

[0060] The unique ID (TUID) assigned to the file is then matched 360 against a list 370 of TUIDs populated through the submission of Title/Artist pairs 370 by labels, publishers, and content owners of those files allowed in the system. In one embodiment, if the TUID is present in the database, the file is allowed to be shared, but if the TUID is not present in the database, the file is blocked. In another embodiment, if the TUID is present in the database, the file is blocked. Either of these embodiments could be applied to files recognized as they are accessed by a user, or transmitted from one computer to another.

[0061] As illustrated in Fig. 4A, an embodiment of the present invention uses a plurality of related databases. Master metadata database 410 contains information on title, artist/author name, owner name and date. Related databases include audio fingerprint database 430 and video fingerprint database 440 which form fingerprint database 290 (Fig. 2). Also included are track length/TOC database 450, text database 460, and hash ID database 470 and guaranteed unique ID database 480.

[0062] As illustrated in Fig. 4B, when unidentified (unknown) recording 100 is accessed by client device 110, information is extracted, including fingerprints 540, 550, metadata 560 and unique ID 570, if present. In addition, the duration 580 of the recording is determined and a numerical hash 590 is calculated. The extracted fingerprints are compared with fingerprints 600, 610. Similarly, matching 620, 630, 640 is performed on the numerical hash, text and unique ID. If a reference recording is located, validation is performed by comparing the duration of unidentified recording 100 with the duration of the reference recording. Results 660-710 with a level of confidence for each method of comparison is supplied to result aggregator 730.

[0063] If no reference recording is found 750 matching unidentified recording 100, the extracted information 540-590 and results are stored in unrecognized holding bin 760 for periodic resubmission to recognition server 120 (Figs. 2 & 4B). In this embodiment, if a reference recording is located 770 with a low aggregate confidence level, post recognition processing 780 is performed by applying heuristics 790, or a manual review 810, e.g., by presenting one or more possible matches to the user and receiving the user's selection in response. The results of such user selections may be included in the heuristics stored in heuristics database 820. If post recognition processing 780 results in identification of a single reference recording or result

aggregator 730 outputs recognized results 770 with a high aggregate confidence level, the hash ID is generated 810 and sent to hash database 480 and client device 110, so that the hash and unique ID (TUID) can be stored in the ID3 tag, if a file is being created.

[0064] In one embodiment, the system learns by watching errors in repeated attempts at recognition of similar files to improve its results. It also may receive manual stimulus from users who indicate that there are errors in the results. This allows recognition to be continuously validated over time. For example a file could be recognized by a system according to the invention, then over time the system determines that recognition of that file was flawed, and indicates to an operator that there was something wrong. In another embodiment, the system determines what is wrong by monitoring non-fingerprint based data and changing the recognition results accordingly.

[0065] The present invention can be utilized to identify any audio content for tracking purposes. Digital audio streams, analog inputs or local audio files, can all be tracked. Such a tracking system could be a server side tracking system deployed at the point of audio delivery and integrated with a reporting, digital rights management (DRM) system, or rights payment system. If the audio content being tracked was from a non-participating third party a client version of the system may be deployed to monitor the content being distributed. In either case, multiple identification methods would be utilized to ensure the highest rate of accuracy.

[0066] Utilizing waveform recognition as a digital rights management component is possible, and can be deployed to compare user created digital audio files with lists of approved content. This enables a filter-in approach within a peer-to-peer file sharing architecture such as the one described above.

[0067] Audio fingerprinting technologies can be used as an anti-piracy tool, and can be customized to the type of audio being investigated. In the case of pirated CDs, the Gracenote's CDDB CD service may be utilized to provide table of content (TOC) recognition to augment audio fingerprinting technologies.

[0068] Identification is the enabling component to deliver value-added services. Without explicit knowledge of the content being distributed it is impossible to distribute value-added content and services that relates to that audio content.

[0069] The many features and advantages of the invention are apparent from the detailed specification and, thus, it is intended by the appended claims to cover all such features and

advantages of the invention that fall within the true spirit and scope of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation illustrated and described, and accordingly all suitable modifications and equivalents may be resorted to, falling within the scope of the invention. For example the system and method have been described as using a unique identifier. However, a hashed identifier could be used instead.

CLAIMS

What is claimed is:

1. A method of identifying recordings, comprising:

extracting information about an unknown recording stored in media possessed by a user and at least one algorithmically determined fingerprint from at least one portion of the unknown recording;

determining a possible identification of the unknown recording using at least one piece of the information extracted from the unknown recording and an identification database of corresponding information for reference recordings; and

identifying the unknown recording when the possible identification based on each of the at least one piece of the information in combination with the at least one algorithmically determined fingerprint identifies a single reference recording with respective confidence levels.

2. A method as recited in claim 1,

wherein the identification database is maintained by a provider of identification services, and

wherein said determining uses a unique identifier from the provider of identification services when the unique identifier is associated with the unknown recording and otherwise uses text associated with the unknown recording when text is associated with the unknown recording.

3. A method as recited in claim 2, further comprising validating said identifying by comparing an extracted length of the unknown recording with a stored length of the single reference recording.

4. A method as recited in claim 3, wherein the text associated with the unknown recording includes a filename of the recording.

5. A method as recited in claim 3, wherein the text associated with the unknown recording includes an ID3 tag for the recording.

6. A method as recited in claim 1, wherein the at least one algorithmically determined fingerprint is extracted from at least one of audio and video information in the at least one portion of the unknown recording.

7. A method as recited in claim 1,

wherein the at least one algorithmically determined fingerprint includes at least two fingerprints, and

wherein said identifying requires each of the at least two fingerprints to identify the single reference recording with respective confidence levels.

8. A method as recited in claim 1, further comprising validating said identifying by comparing an extracted length of the unknown recording with a stored length of the single reference recording.

9. A method as recited in claim 8, further comprising:

repeating said extracting, determining and identifying for a plurality unknown recordings from a plurality users;

monitoring unsuccessful identifications of the unknown recordings; and

detecting a possible error in the identification database from a pattern of errors.

10. A method as recited in claim 9, wherein said monitoring includes receiving information from the users indicating that said identifying was incorrect.

11. A method as recited in claim 9,

wherein said monitoring includes storing the at least one algorithmically determined fingerprint and an identifier of the single reference recording when said validating is not successful, and

wherein said method further comprises indicating the possible error in the identification database when substantially different fingerprints are stored for a single identifier.

12. A method as recited in claim 9, wherein said method further comprises indicating the possible error when the at least one algorithmically determined fingerprint matches one of the reference recordings, but the unknown recording is associated with ID3 Tag information different from that of the one of the reference recordings.

13. A method as recited in claim 9, further comprising correcting the possible error based on the information extracted from the unknown recordings.

14. A method as recited in claim 8, further comprising indicating a possible error in the identification database when the at least one algorithmically determined fingerprint is substantially similar to one of the reference recordings, but substantially different information is extracted from the unknown recording.

15. A method as recited in claim 1, further comprising delivering related data for the unknown recording from a supplemental database to supplement data embedded within the unknown recording for display and user manipulation.

16. A method of identifying recordings, comprising:

extracting fingerprints from at least one portion of an unknown recording using a plurality of algorithms;

determining a possible identification of the unknown recording using at least two of the fingerprints extracted from the unknown recording and at least one database of correspondingly generated fingerprints for reference recordings; and

identifying the unknown recording when the possible identification based on each of the fingerprints identifies a single reference recording with respective confidence levels.

17. A method as recited in claim 16, wherein each fingerprint is extracted from at least one of audio and video information in the at least one portion of the unknown recording.

18. A method as recited in claim 16, further comprising validating said identifying by comparing a length of the unknown recording with a stored length of the single reference recording.

19. A method as recited in claim 18,

wherein said extracting is performed by client equipment possessed by a user,

wherein said determining, identifying and validating are performed by at least one server under control of a provider of identification services, and

wherein said method further comprises:

transmitting a unique identifier associated with the single reference recording from the at least one server to the client equipment after said validating is successful; and

associating the unique identifier with the unknown recording in the client equipment.

20. A method as recited in claim 19, further comprising:

comparing the unique identifier with a permission list of stored identifiers in the at least one database;

indicating that the recording may be shared if there is a match for the unique identifier in the permission list.

21. A method as recited in claim 19, further comprising:

comparing the unique identifier with a block list of stored identifiers in the at least one database;

indicating that the recording may not be shared if there is a match for the unique identifier in the block list.

22. A method of obtaining reference information stored in a database used to identify unknown recordings, comprising:

obtaining non-waveform data associated with a recording possessed by a user of the database for identification of recordings possessed by the user;

extracting at least one fingerprint from at least one portion of the recording; and

storing the at least one fingerprint as identifying information for the recording, when a match is found in the database for the non-waveform data.

23. A method as recited in claim 22, wherein the non-waveform data indicates the length of the recording.

24. A method as recited in claim 23, wherein the recording is permanently stored on a removable medium and the non-waveform data is derived from table of contents data for the recording.

25. A method as recited in claim 22, wherein the non-waveform data includes text associated with the recording.

26. A method as recited in claim 25, wherein the non-waveform data includes and ID3 tag.

27. A method as recited in claim 26,

wherein the ID3 tag includes encoded information,

wherein the database is maintained by a provider of identification services and the encoded information is generated under control of the provider of identification services, and

wherein said method further comprises validating the non-waveform data by decoding the encoded information prior to said storing of the identifying information in the database.

28. A method as recited in claim 25, wherein the non-waveform data includes a watermark.

29. A method as recited in claim 25, wherein the non-waveform data includes media information regarding source media type.

30. A method as recited in claim 29, wherein the media information identifies the source media type as CD-R.

31. A method as recited in claim 29, wherein the media information identifies the source media type as CD-DA.

32. A method as recited in claim 29, wherein the media information identifies the source media type as a digital file.

33. A method as recited in claim 29, wherein the media information identifies the source media type as a digital versatile disc.

34. A method as recited in claim 25, wherein the text includes a filename of the recording.

35. A method as recited in claim 25, wherein the text includes a title of the recording.

36. A method as recited in claim 25, wherein the text includes an artist name of a participant in creation of the recording.

37. A method as recited in claim 25, wherein the text includes an album name associated with the recording.

38. A method as recited in claim 22, wherein said obtaining and extracting are performed by client equipment possessed by a plurality users for different copies of the recording and different users extract different fingerprints from the recording.

39. A method as recited in claim 38, further comprising:

maintaining the database on at least one server under control of a provider of identification services, and

transmitting from the at least one server to the client equipment, extraction instructions on which of the different fingerprints each of the client equipment extracts.

40. A method as recited in claim 39, further comprising:

transmitting the non-waveform data from the client equipment to the at least one server; and

selecting the extraction instructions by the at least one server for said transmitting to the client equipment based on the non-waveform data.

41. A method as recited in claim 40, further comprising updating the extraction instructions based at least in part on frequency of receipt of the non-waveform data for the recording.

42. A method as recited in claim 40, wherein said selecting of the extraction instructions is based at least in part on type of the client equipment receiving the extraction instructions.

43. A method as recited in claim 40, wherein said selecting of the extraction instructions is based at least in part on geographical location of the client equipment receiving the extraction instructions.

44. A method as recited in claim 40, wherein said selecting of the extraction instructions is based at least in part on software operating on the client equipment receiving the extraction instructions.

45. A method as recited in claim 40, further comprising updating the extraction instructions based at least in part on number of users who have supplied the identifying information.

46. A method as recited in claim 40, wherein said selecting of the extraction instructions is based at least in part on quality of the copies of the recording.

47. A method as recited in claim 40, further comprising transmitting the at least one fingerprint from the client equipment to the at least one server at a time specified by the extraction instructions.

48. A method as recited in claim 47, further comprising storing the at least one fingerprint at the client equipment until a specified number of fingerprints are ready for said transmitting.

49. A method as recited in claim 47, wherein said transmitting of the at least one fingerprint occurs when a communication channel with the at least server is available.

50. A method as recited in claim 47, wherein said transmitting of the at least one fingerprint for a first recording accessed by a piece of client equipment occurs with said transmitting of the non-waveform data for a second recording accessed by the piece of client equipment.

51. A method as recited in claim 47,

wherein the recording is permanently stored on a removable medium and the client equipment generates at least one encoded file from the recording, and

wherein said transmitting transmits the at least one fingerprint before encoding the recording.

52. A method as recited in claim 47,

wherein the recording is permanently stored on a removable medium and the client equipment generates at least one encoded file from the recording, and

wherein said transmitting transmits the at least one fingerprint after encoding one track of the removable medium.

53. A method as recited in claim 47,

wherein the recording is permanently stored on a removable medium and the client equipment generates at least one encoded file from the recording, and

wherein said transmitting transmits the at least one fingerprint after receiving an indication that encoding of the removable medium has been completed.

54. A method as recited in claim 22, wherein the database includes the identifying information for musical recordings.

55. A method as recited in claim 22, wherein the database includes the identifying information for video recordings.

56. A method as recited in claim 22, further comprising:

detecting a quality of the at least one fingerprint;

identifying another copy of the recording using the at least one fingerprint; and

replacing the at least one fingerprint with a higher quality fingerprint when the other copy of the recording produces the higher quality fingerprint.

57. A method as recited in claim 56, wherein said detecting of the quality is based on an encoding technique used for the recording.

58. A method as recited in claim 56, wherein said detecting of the quality is based on a media type used to store the recording.

59. A method as recited in claim 56, wherein said detecting of the quality is based on error correction capability of user equipment accessing the recording.

60. A method as recited in claim 59, wherein said detecting of the quality assigns higher quality when hardware error correction is used than when software error correction is by the user equipment.

61. A method as recited in claim 56, wherein said detecting of the quality is based on number of errors detected during said extracting of the fingerprint.

62. A method as recited in claim 22,

wherein said obtaining and extracting are performed by client equipment possessed by a plurality users for different copies of the recording, and

wherein said method further comprises:

comparing the at least one fingerprint obtained from one of the users with the at least one fingerprint extracted from at least one other user; and

updating the at least one fingerprint in the database based on said comparing.

63. A method as recited in claim 62, wherein said updating is performed after said comparing determines that fingerprints from different users have a predetermined correlation.

64. A method as recited in claim 62, wherein said updating combines fingerprints from different users for storage in the database.

65. A system for identifying recordings, comprising:

an extraction unit to extract information about an unknown recording stored in media possessed by a user and at least one algorithmically determined fingerprint from at least one portion of the unknown recording; and

an identification unit, coupled to said extraction unit, to make a possible identification of the unknown recording using at least one piece of the information extracted from the unknown recording and an identification database of corresponding information for reference recordings, and to identify the unknown recording when the possible identification based on each of the at least one piece of the information in combination with the at least one algorithmically determined fingerprint identifies a single reference recording with respective confidence levels.

66. A system for identifying recordings, comprising:

an extraction unit to extract fingerprints from at least one portion of an unknown recording using a plurality of algorithms; and

an identification unit, coupled to said extraction unit, to make a possible identification of the unknown recording using at least two of the fingerprints extracted from the unknown recording and at least one database of correspondingly generated fingerprints for reference recordings, and to identify the unknown recording when the possible identification based on each of the fingerprints identifies a single reference recording with respective confidence levels.

67. A system for obtaining reference information stored in a database used to identify unknown recordings, comprising:

a receiving unit to obtain non-waveform data associated with a recording possessed by a user of the database for identification of recordings possessed by the user;

an extraction unit to extract at least one fingerprint from at least one portion of the recording; and

a storage unit, coupled to said receiving unit and said extraction unit, to store the at least one fingerprint as identifying information for the recording, when a match is found in the database for the non-waveform data.

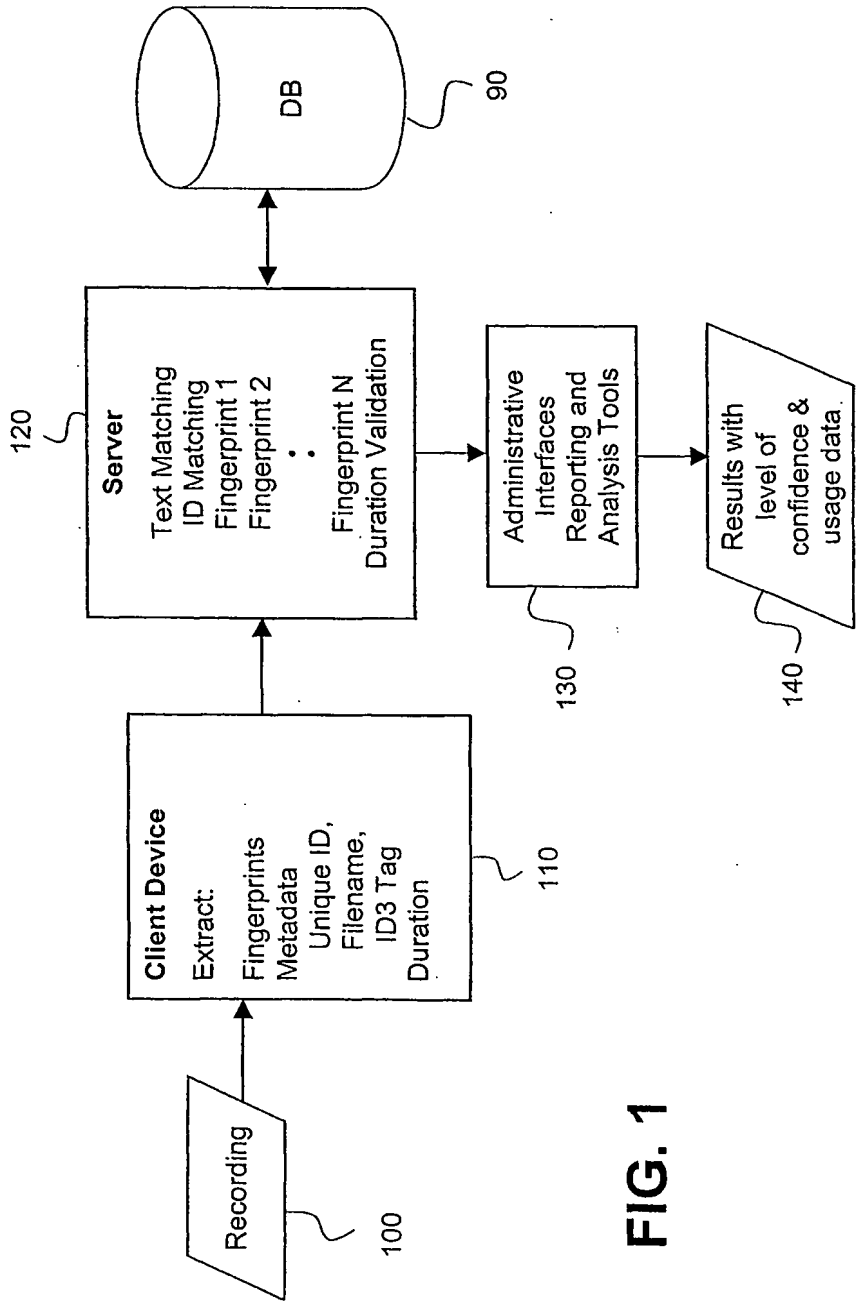
DB
90
Server
Text Matching
ID Matching
Fingerprint 1
Fingerprint 2
:
Fingerprint N
Duration Validation
120
Administrative
Interfaces
Reporting and
Analysis Tools
130
Results with
level of
confidence &
usage data
140
Client Device
Extract:
Fingerprints
Metadata
Unique ID,
Filename,
ID3 Tag
Duration
110
Recording
100

FIG. 1

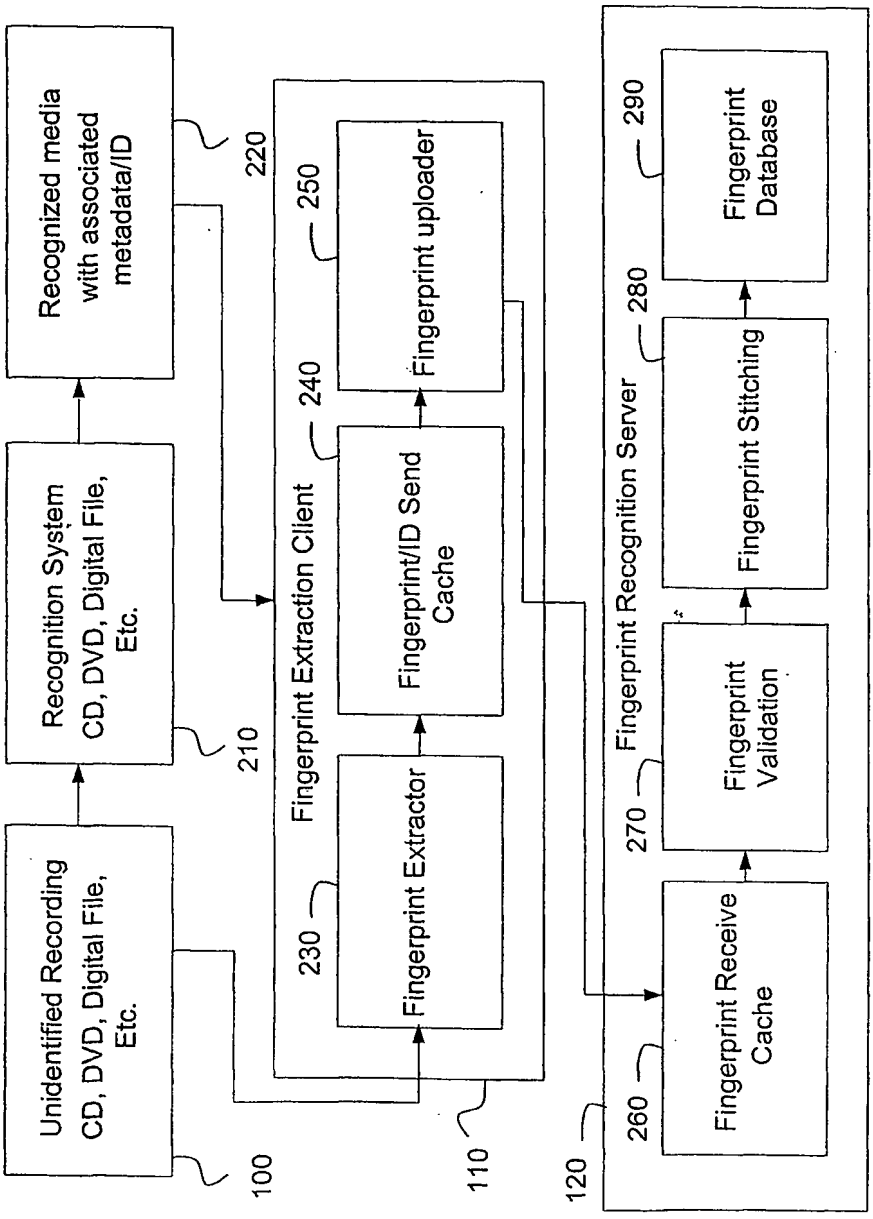
Unidentified Recording
CD, DVD, Digital File, Etc.
100
Recognition System
CD, DVD, Digital File, Etc.
210
Recognized media with associated metadata/ID
220
Fingerprint Extraction Client
230
Fingerprint Extractor
240
Fingerprint/ID Send Cache
250
Fingerprint uploader
110
120
Fingerprint Recognition Server
260
Fingerprint Receive Cache
270
Fingerprint Validation
280
Fingerprint Stitching
290
Fingerprint Database

FIG. 2

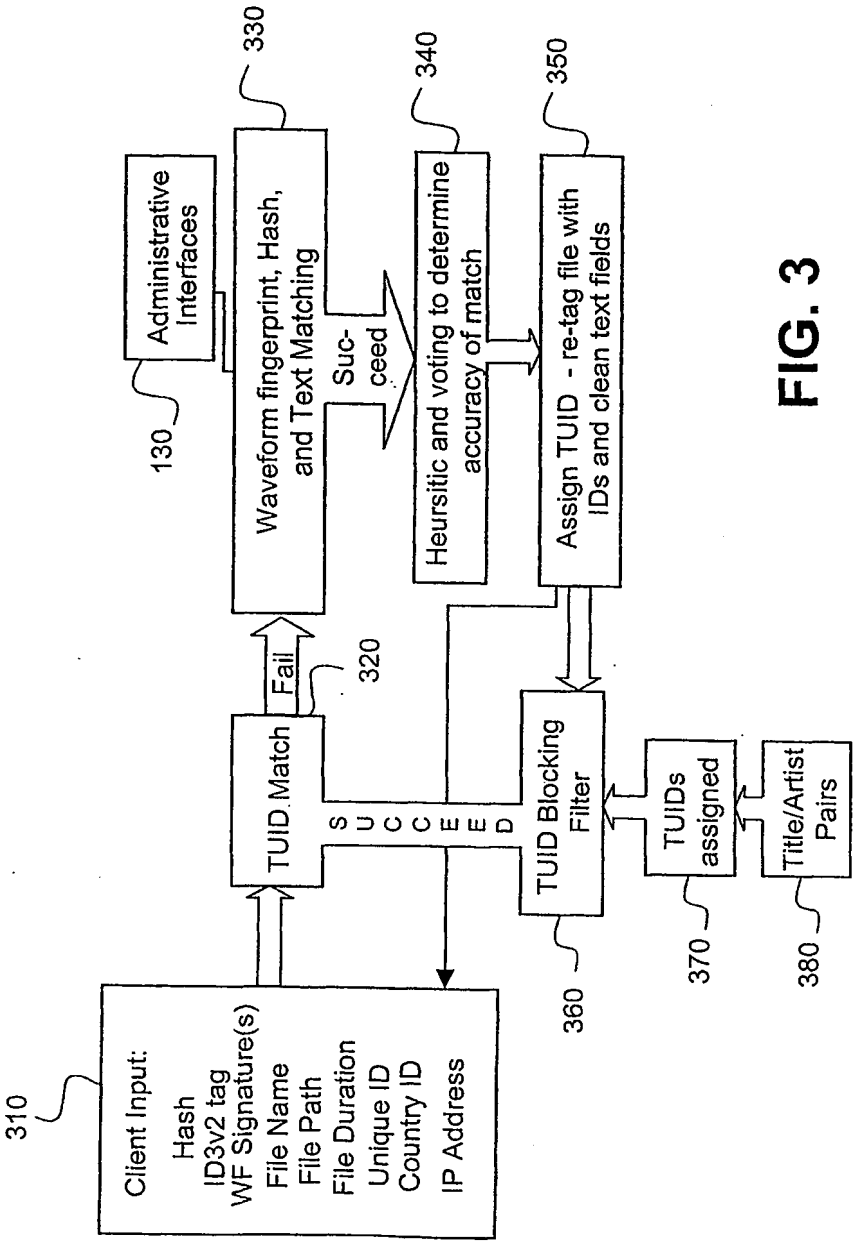
310
Client Input:
Hash
ID3v2 tag
WF Signature(s)
File Name
File Path
File Duration
Unique ID
Country ID
IP Address
TUID Match
Fail
320
SUCCEED
TUID Blocking Filter
360
TUIDs assigned
370
Title/Artist Pairs
380
Assign TUID - re-tag file with IDs and clean text fields
350
Heursitic and voting to determine accuracy of match
340
Waveform fingerprint, Hash, and Text Matching
330
Administrative Interfaces
130

FIG. 3

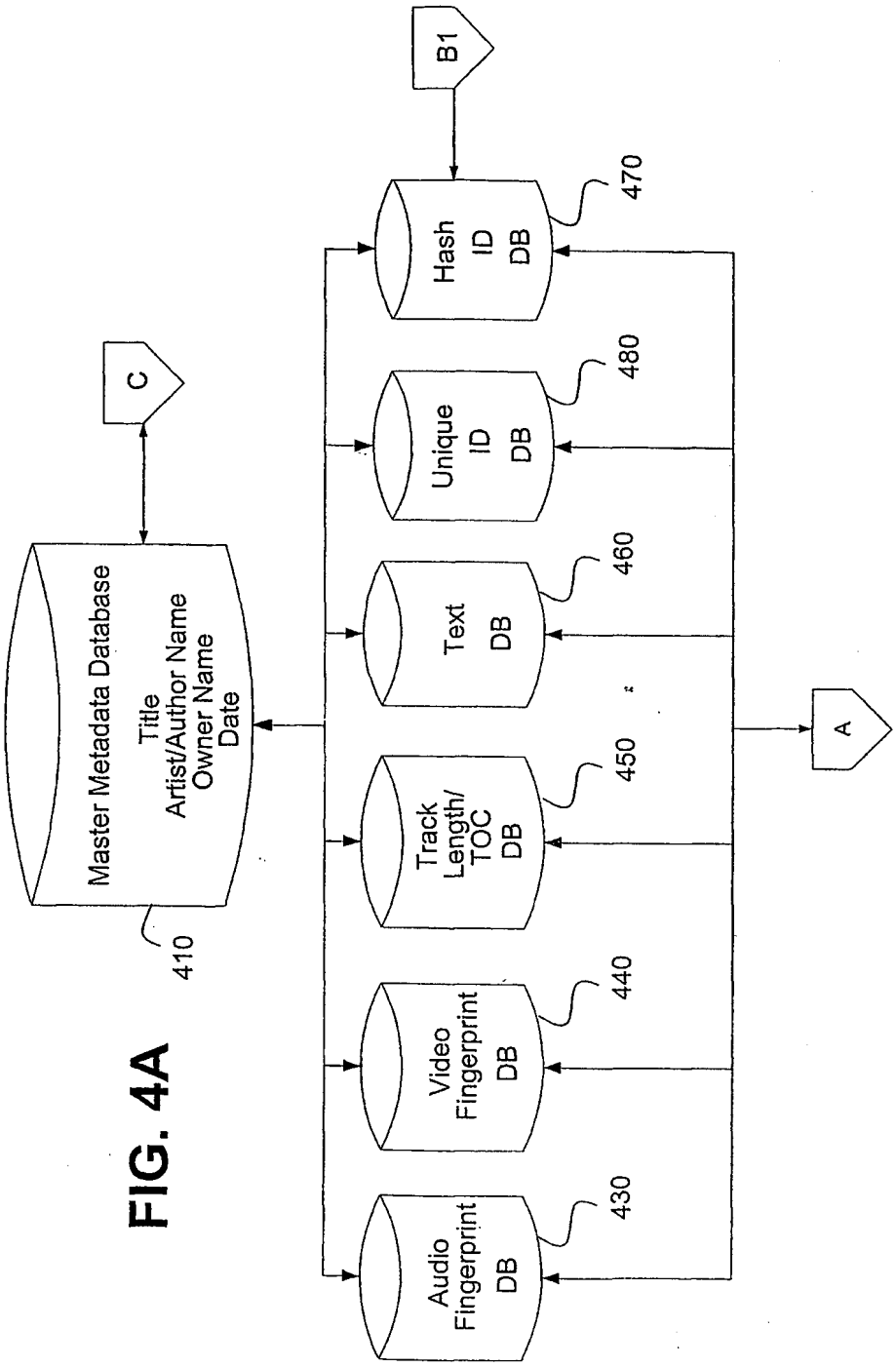
B1
Hash ID DB
470
C
Unique ID DB
480
Master Metadata Database
Title
Artist/Author Name
Owner Name
Date
Text DB
460
A
Track Length/ TOC DB
450
Video Fingerprint DB
440
Audio Fingerprint DB
430
410

FIG. 4A

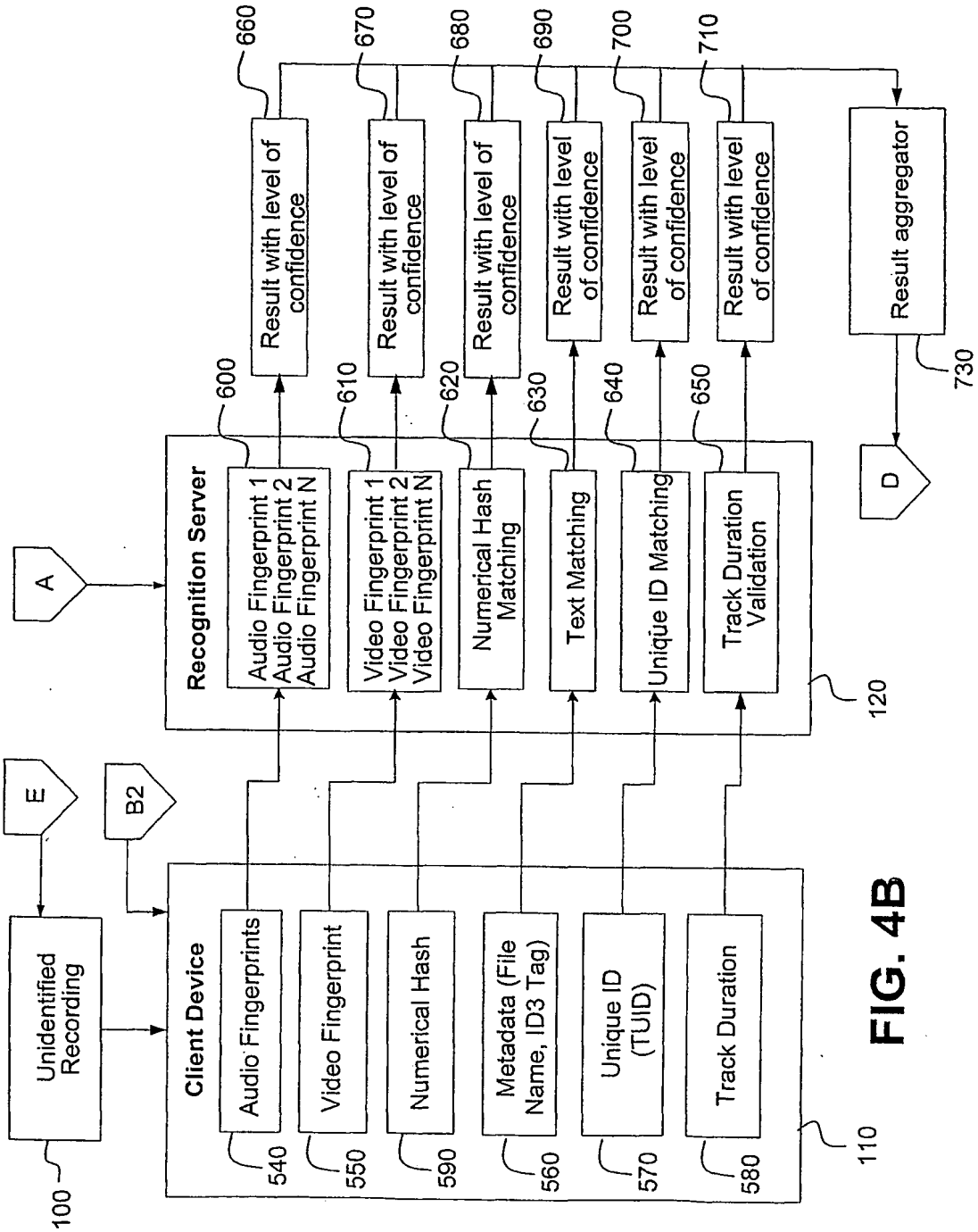
Unidentified Recording
100
Client Device
Audio Fingerprints
540
Video Fingerprint
550
Numerical Hash
590
Metadata (File Name, ID3 Tag)
560
Unique ID (TUID)
570
Track Duration
580
110
E
B2
A
Recognition Server
Audio Fingerprint 1
Audio Fingerprint 2
Audio Fingerprint N
600
Video Fingerprint 1
Video Fingerprint 2
Video Fingerprint N
610
Numerical Hash Matching
620
Text Matching
630
Unique ID Matching
640
Track Duration Validation
650
120
Result with level of confidence
660
Result with level of confidence
670
Result with level of confidence
680
Result with level of confidence
690
Result with level of confidence
700
Result with level of confidence
710
Result aggregator
730
D

FIG. 4B

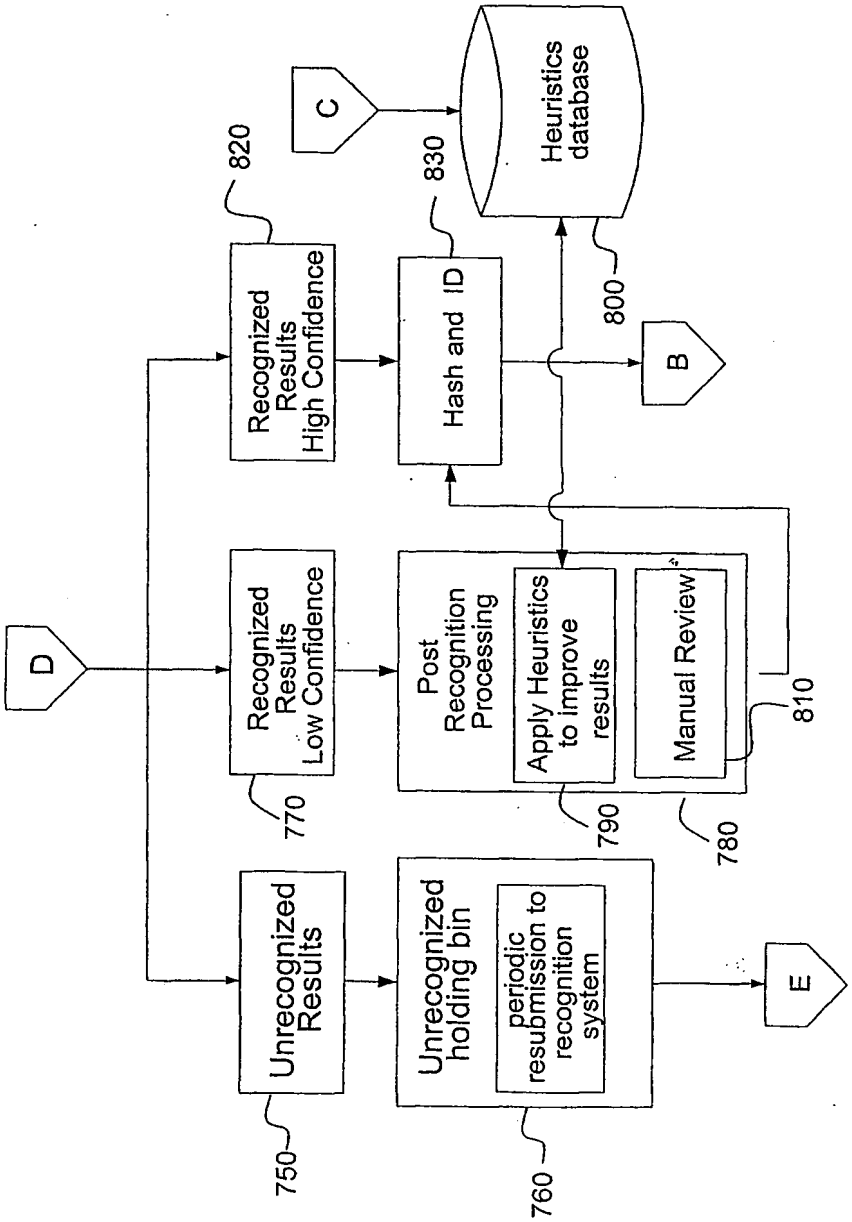
D
Recognized Results Low Confidence
770
Post Recognition Processing
Apply Heuristics to improve results
Manual Review
790
780
810
Recognized Results High Confidence
820
Hash and ID
830
C
Heuristics database
800
B
Unrecognized Results
750
Unrecognized holding bin
periodic resubmission to recognition system
760
E

FIG. 4C